WHAT IS CLAIMED IS:

1. A computer-implemented method of extracting information from an information source, comprising:

accessing strings in the information source; and

comparing the strings in the information source with generalized extraction patterns and identifying strings in the information source that match at least one generalized extraction pattern, the generalized extraction patterns including words and wildcards, wherein the wildcards denote that at least one word in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern.

2. The computer-implemented method of claim 1 and further comprising extracting at least two elements from strings in the information source that have been identified to match, the at least two elements being based on at least two corresponding elements in a corresponding generalized extraction patterns.

3. The computer-implemented method of claim 2 wherein for at least one of the corresponding elements in each of the generalized extraction patterns, there is at least one word positioned between said at least one of the corresponding elements and the wildcards.

4. The computer-implemented method of claim 1 wherein the wildcards indicate the number of words that can be skipped.

5. A computer-readable medium for extracting information from an information source, comprising:

a data structure including a set of generalized extraction patterns including words and an indication of a position for at least one optional word; and

an extraction module using the set of generalized extraction patterns to match strings in the information source with the generalized extraction patterns.

6. The computer-readable medium of claim 5 wherein the generalized extraction patterns further include at least two elements related to a subject.

7. The computer-readable medium of claim 6 wherein for the generalized extraction patterns there is at least one word positioned between at least one of the elements and the indication.

8. The computer-readable medium of claim 5 wherein the indication includes a number of words that can be skipped during information extraction.

9. A method of generating patterns for use in extracting information from an information source, comprising:

establishing a set of strings including at least two elements corresponding to a subject;

generating a set of generalized extraction patterns that correspond to the set of strings, the generalized extraction patterns including the at least two elements, words and an indication of a position for at least one optional word.

10. The method of claim 9 and further comprising removing patterns from the set of generalized extraction patterns that do not meet a frequency threshold in the set of strings.

11. The method of claim 9 and further comprising removing patterns from the set of generalized extraction patterns that contain the indication adjacent to one of the at least two elements in the generalized extraction pattern.

12. The method of claim 9 and further comprising removing patterns from the set of generalized extraction patterns where the number of words to be skipped by the indication is above a threshold.

13. The method of claim 9 and further comprising ranking the generalized extraction patterns in the set of generalized extraction patterns.

14. The method of claim 13 wherein the step of ranking further comprises calculating a precision score for each generalized extraction pattern.

15. The method of claim 13 and further comprising removing patterns from the set of generalized extraction patterns that do not meet a ranking threshold.

16. The method of claim 9 and further comprising determining a number of words that a particular indication will skip.

17. A method of generating patterns for use in extracting information from an information source, comprising:

establishing a set of strings including at least two elements corresponding to a subject;

identifying consecutive patterns within the set of strings that include words and the at least two elements; and

generating a set of generalized extraction patterns from the consecutive patterns identified, the generalized extraction patterns including the at least two elements, words and wildcards, wherein the

wildcards express a combination of the consecutive patterns.

18. The method of claim 17 and further comprising removing patterns from the set of generalized extraction patterns that do not meet a frequency threshold in the set of strings.

19. The method of claim 17 and further comprising removing patterns from the set of generalized extraction patterns that contain a wildcard adjacent to one of the at least two elements in the generalized extraction pattern.

20. The method of claim 17 and further comprising removing patterns from the set of generalized extraction patterns where the number of words to be skipped by a wildcard is above a threshold.

21. The method of claim 17 and further comprising ranking the generalized extraction patterns in the set of generalized extraction patterns.

22. The method of claim 21 wherein the step of ranking further comprises calculating a precision score for each generalized extraction pattern.

23. The method of claim 21 and further comprising removing patterns from the set of generalized

extraction patterns that do not meet a ranking threshold.

24. The method of claim 17 and further comprising determining a number of words that a particular wildcard will skip.